SCIENCE · PASSION · TECHNOLOGY

TU Graz

Open Thesis / Project

Parameter-efficient Fine-tuning Framework for Deep Models on IoT Devices

Embedded Learning and Sensing Systems Group

Motivation

In deep learning, a core training strategy includes extensive pre-training on a large-scale dataset followed by its fine-tuning on a specific task. As deep models grow in size, the approach of full fine-tuning, *i.e.*, updating all the weights, becomes progressively more challenging, *i.e.*, resource- and data-hungry. Parameter-efficient fine-tuning (PEFT) methods seek to address this challenge by updating only a small subset of model parameter. Existing literature predominantly focuses on the applications of PEFT methods within the domain of natural language processing for large language models. The goal of this project is to evaluate the performance and cost implications of PEFT, in terms of memory size and computation, when applying these methods to deep networks designed to run on highly resource-constrained IoT devices. You will build a framework that allows using PEFT with TFL Micro for model adaptation on IoT devices.

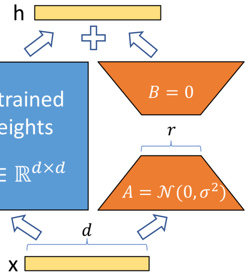
Interested? Please contact us for more details!

Target Group

Students in ICE, Computer Science, Software Eng.

Thesis Type

Master Project / Master Thesis.


h
Pretrained Weights
$W \in \mathbb{R}^{d \times d}$
$B = 0$
r
$A = \mathcal{N}(0, \sigma^2)$
d
x

LoRA proposed reparametrization

Goals and Tasks

The goal of this work is to evaluate and compare the performance of PEFT methods when fine-tuning pretrained models (*e.g.*, MobileNet) on new datasets. The project includes the following tasks:

- In-depth understanding of neural networks standard training using backpropagation;
- Literature research on existing PEFT methods and available implementations;
- Implement / port and compare PEFT methods, evaluate PEFT performance and computational costs;
- Summarize the results in a written report.

Requirements:

- Interest to explore and analyze the performance of PEFT methods;
- Programming skills in Python;
- Prior experience of machine learning frameworks (*e.g.*, Tensorflow, Pytorch).

Used Tools & Equipment

- A laptop (GPU infrastructure will be provided if needed).

Contact Persons

- Francesco Corti (francesco.corti@tugraz.at)
- Assoc. Prof. Olga Saukh (saukh@tugraz.at)

Institute for Technical Informatics
Embedded Learning and Sensing Systems Group